Genomics **71,** 150–155 (2001)
doi:10.1006/geno.2000.6433, available online at http://www.idealibrary.com on IDEAL®

Structure of the Mouse Stat 3/5 Locus: Evolution from *Drosophila* to Zebrafish to Mouse

Keiko Miyoshi,* Yongzhi Cui,* Greg Riedlinger,* Phyllis Robinson, Jessica Lehoczky,† Leonard Zon,‡ Takami Oka,* Ken Dewar,† and Lothar Hennighausen*,[1]

*Laboratory of Genetics and Physiology, National Institute of Diabetes, Digestive and Kidney Diseases, National Institutes of Health, 9000 Rockville Pike, Building 8, Room 101, Bethesda, Maryland 20892; †Whitehead Institute, Massachusetts Institute of Technology, One Cambridge Square, Cambridge, Massachusetts 02142; and ‡Children's Hospital, Harvard Medical School, Boston, Massachusetts 02115

Received October 23, 2000; accepted November 2, 2000

Signal transducers and activators of transcription (Stat) are transcription factors that can be activated by many cytokines. While *Drosophila* contains only one Stat (d-Stat), mammals contain seven, with STATs 3, 5a, and 5b being the closest functional relatives. To understand the evolutionary relationship between d-Stat and vertebrate STATs 3 and 5, we isolated, sequenced, and analyzed the zebrafish Stat3 (z-Stat3) gene and a 500-kb region spanning mouse chromosome 11, 60.5 cM containing three Stat genes (m-Stats). Within this region we identified the genes encoding m-Stats 3, 5a, and 5b, Cnp1, Hcrt/Orexin, Ptrf, GCN5, mDj11, and four new genes. The 5′ ends of the m-Stat5a and m-Stat5b genes are juxtaposed to each other, and the 3′ ends of the m-Stat3 and Stat5a genes face each other. While the m-Stat5a and m-Stat3 genes have one promoter each, which are active in many tissues, the m-Stat5b gene acquired two distinct promoters. The distal promoter is expressed ubiquitously, and transcription from the proximal promoter is restricted to liver, muscle, and mammary tissue. Through a comparison of exon–intron boundaries from the m-Stat3, m-Stat5a, and m-Stat5b, z-Stat3, and d-Stat genes, we deduced their evolutionary relationship. We propose that the Stat3 and Stat5 lineages are derived from the duplication of a common primordial gene and that d-Stat is a part of the Stat5 lineage.

INTRODUCTION

Signal transducers and activators of transcription (STATs) are transcription factors that mediate cytokine signaling from the cell surface receptor to the nucleus (Takeda and Akira, 2000; Leonard and O'Shea, 1998; Chatterjee-Kishore *et al.*, 2000). Mammals contain seven Stats (Stats 1, 2, 3, 4, 5a, 5b and 6), each of them serving distinct cytokines. The genes encoding mouse STATs 1 and 4 are located on chromosome 1, STATs 2 and 6 are on chromosome 10, and STATs 3, 5a, and 5b (m-Stats) are on chromosome 11. In contrast, *Drosophila* contains only one Stat gene (Hou *et al.*, 1996). No STATs have been identified in yeast, suggesting that these cytokine-signaling pathways are unique to multicellular organisms. Among the seven mouse Stats, Stats 5a and 5b are the closest relatives to the *Drosophila* homologue d-Stat and share functional roles in that both control cell survival and differentiation in the hematopoietic system (Hanratty and Dearolf, 1993; Harrison *et al.*, 1995). Stat 1 and Stat 3 (z-Stat3) have been identified in zebrafish (Oates *et al.*, 1999). The Stat3 protein is very conserved between zebrafish and mouse but gene expression appears to be conserved only in the central nervous system. Mouse Stat5a is critical for the survival and differentiation of mammary epithelial cells (Liu *et al.*, 1997; Humphreys and Hennighausen, 1999), and Stat 3 controls cell death in mammary epithelium during involution (Chapman *et al.*, 1999).

Although m-Stat3, m-Stat5, and d-Stat contain conserved domains, their evolutionary origin and relationship are not clear. To establish the evolutionary trail, we isolated and characterized mouse Stat3, Stat5a, and Stat5b as well as zebrafish Stat3. We propose the evolutionary relationship among these genes and distant species through the comparison of the exon–intron boundaries.

MATERIALS AND METHODS

Isolation and DNA sequence analysis of the m-Stat 3/5 locus. Two phage λ clones containing sequences from the mouse Stat5a and Stat5b genes (129 SVJ) were identified and sequenced (GenBank Accession Nos. AF049104 and AF234171). Four BAC clones containing the genes encoding m-STATs 3, 5a, and 5b were identified on mouse ES (release I) genomic filters (Genome Systems Inc.) by screening with PCR-generated probes specific for STATs 3, 5a, and 5b. The Stat3-specific probe was generated with primers 5′-TGG-AGA-GGT-AAC-AGC-CCC-TTG-TAG-3′ and 5′-GAT-TGG-TCA-GCT-CAC-AGA-AAT-GC-3′ and spanned 880 bp. The Stat5a-specific probe was generated with primers 5′-GTG-ATT-TTG-TTT-CAC-CTG-CCT-AGC-3′ and 5′-TCG-GAC-AGT-GCT-TCT-TCT-CTT-CC-3′ and spanned 932 bp. The Stat5b-specific probe was generated with primers 5′-GTT-TGG-AAA-AGC-TTG-TTG-GTG-G-3′ and 5′-TCT-GAA-GGT-TTT-CAG-AAC-ACC-3′ and spanned 347 bp. The

[1] To whom correspondence should be addressed. Telephone: (301) 496-2716. Fax: (301) 480-7312. E-mail: mammary@nih.gov.

0888-7543/01 $35.00

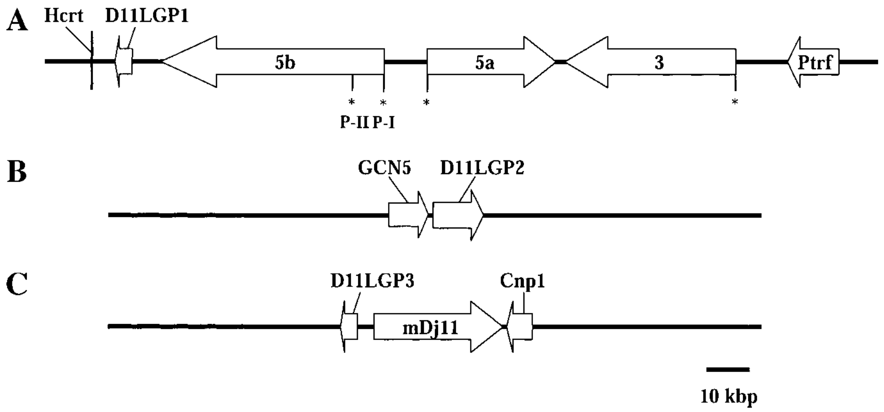

FIG. 1. Genes in the mouse Stat3/5 locus. The BAC and λ clones span approximately 500 kb and encompass eight known genes and four new genes. The Ptrf gene upstream of the Stat3 gene delineates one end of the sequenced region. Since two gaps remained in the 500-kb sequence, the orientation of two contigs with respect to the Stat genes remains to be determined. Four new genes (D11LGP1–4) were identified downstream of Stat5b. (**A**) The m-Stat5a gene is 30 kb, the m-Stat5b gene is more than 54 kb, and the m-Stat3 gene is more than 37 kb. The asterisks indicate the start sites of transcription (note that Stat5b has two promoters, PI and PII). The Hcrt and D11LGP1 genes are downstream of the Stat5b gene, and the Ptrf gene is upstream of the Stat3 gene. (**B**) The GCN5 and D11LGP2 genes are downstream of the Stat5b gene. (**C**) The Cnp1, mDj11, and D11LGP3 genes are also downstream of the Stat5b gene. The gene sizes and intergenic sequences are drawn to scale.

filters were screened with [α-^{32}P]dCTP-labeled probes in QuickHyb hybridization solution (Stratagene) at 68°C overnight. Four positive clones were further characterized using PCR. As these clones contained overlapping regions of the genome, two of them were subjected to sequence analysis. Clone L5902 (GenBank Accession No. AC021632) contained the Stat5b gene, and clone L5725 (GenBank Accession No. AC023193) contained the Stat3 gene and the 3′ part of the Stat5a gene. The 5′ part and the promoter of the Stat5a gene were obtained from a third BAC clone and a phage λ clone (GenBank Accession No. AF049104).

The zebrafish Stat3 gene was isolated from a BAC clone and sequenced (GenBank Accession No. AF322857). Alignment of mRNA sequences with genomic sequences was achieved with the Blast 2 sequence software from GenBank. The search for known and unknown genes in the BAC clones was performed using GenBank's nr and dbEST search modes, respectively.

The position of introns. The positions of introns with respect to the amino acids within d-Stat, m-Stat5a, m-Stat5b, m-Stat3, and z-Stat3 have been determined. The introns in the d-Stat gene coincide with the following amino acid positions: 43, 386, 507, 632, 699, and 745. The introns in the m-Stat5a gene coincide with the following amino acid positions: 44, 95, 126, 184, 228, 279, 330, 390, 420, 461, 492, 561, 593, 636, 688, 706, and 740. The introns in the m-Stat5b gene coincide with the following amino acid positions: 44, 95, 126, 184, 228, 279, 330, 390, 420, 461, 492, 561, 593, 636, 693, 711, and 746. The introns in the m-Stat3 gene coincide with the following amino acid positions: 44, 92, 125, 154, 185, 216, 267, 320, 351, 371, 381, 412, 428, 456, 489, 535, 552, 584, 631, 702, 716, and 754. The introns in the z-Stat3 gene coincide with the following amino acid positions: 44, 125, 154, 183, 217, 268, 321, 352, 372, 381, 413, 429, 457, 490, 535, 553, 585, 631, 704, 719, 739, and 772.

RNA blot hybridization. Mouse Multiple Tissue Northern Blot membranes (MTN, Clontech), which contain approximately 2 μg of poly(A)$^+$ RNA per lane from eight different tissues, were hybridized with random-primed [α-^{32}P]dCTP-labeled probes in ExpressHyb solution for 1 h at 65°C. Washes were performed in 0.5× SSC, 0.1% SDS at 65°C. The membranes were exposed to X-ray films. PCR fragments were used as hybridization probes. The Stat3 3′UTR-specific probe was generated with primers 5′-TGC-TAT-CTT-TGG-GCA-ATC-TGG-3′ and 5′-AAC-CTC-CTG-GGC-TTA-GTC-C-3′ and spanned 325 bp. The Stat5a 3′UTR-specific probe was generated with primers 5′-CGT-TCC-CAC-CAT-CCC-TTT-TC-3′ and 5′-TAC-AGG-TTT-TTC-GCC-CCA-GG-3′ and spanned 304 bp. The Stat5b 5′UTR (detected transcripts from promoter II)-specific probe was generated with primers 5′-AGA-CAG-AGA-AGA-CAG-TGG-AG-3′ and 5′-AAG-CCG-TTA-GAA-GCA-GGA-GC-3′ and spanned 457 bp. The Stat5b 3′UTR (detected transcripts from promoter I and II)-specific probe was generated with primers 5′-ATG-CGA-CTG-TCC-CAG-TAA-CC-3′ and 5′-CCA-CTT-CAA-ATC-TCC-TTC-CAC-3′ and spanned 408 bp. The Hcrt-specific probe was generated with primers 5′-TGG-GGT-GGA-CGC-ACA-GCC-3′ and 5′-AGG-ACA-AGG-ATA-GAA-GAT-GGG-3′ and spanned 300 bp. The Cnp1-specific probe was generated with primers 5′-TAT-TTT-ACA-ACA-GGT-GAA-GGG-G-3′ and 5′-TTA-CAG-GTA-CGT-TAG-CTT-TGA-G-3′ and spanned 323 bp. The Ptrf-specific probe was generated with primers 5′-AGA-TCA-TGA-CTT-TGA-AGT-TGC-G-3′ and 5′-GGA-TGT-CAC-GCT-CCA-TAT-CG-3′ and spanned 467 bp. The GCN5-specific probe was generated with primers 5′-GGA-CAC-AGA-CAC-CAA-ACA-AG-3′ and 5′-GGG-AAG-TGA-GTG-AGG-ATG-AG-3′ and spanned 500 bp. The mDj11-specific probe was generated with primers 5′-GGC-TAA-TGA-AGG-CCA-ACT-AC-3′ and 5′-ATT-CAC-AAG-CTG-CCT-GCT-TC-3′ and spanned 219 bp.

RESULTS

The Mouse Stat3/5 Locus—Structure and Genes

We isolated and sequenced two phage λ clones, several cDNA clones, and two BAC clones spanning a total of 500 kb of the chromosomal locus-containing mouse STATs 3, 5a, and 5b (Fig. 1A). The promoters of the Stat 5a and 5b genes are located head-to-head and are separated by 10 kb. The Stat3 gene is located next to the Stat5a gene, and their polyadenylation sites are 3 kb apart (tail-to-tail). Furthermore, within this 500 kb of sequence, we identified five additional known genes and four new genes. The gene encoding the polymerase I transcription releasing factor (Ptrf) is located more than 14 kb upstream of the Stat 3 gene, and the gene encoding hypocretin/orexin (Hcrt) is located more than 18 kb downstream of the Stat5b gene (Fig. 1A). A new gene, called D11LGP1 (GenBank Accession Nos. AF316996 and AF316998), is located between Hcrt and Stat5b. A second new gene, called D11LGP2 (GenBank Accession Nos. AF316999 and AF317000), and the gene encoding GCN5 are located downstream of Hcrt (Fig. 1B). The Cnp1 gene encoding 2′-, 3′-cyclic-nucleotide 3′-phosphodiesterase and the mouse DnaJ homo-

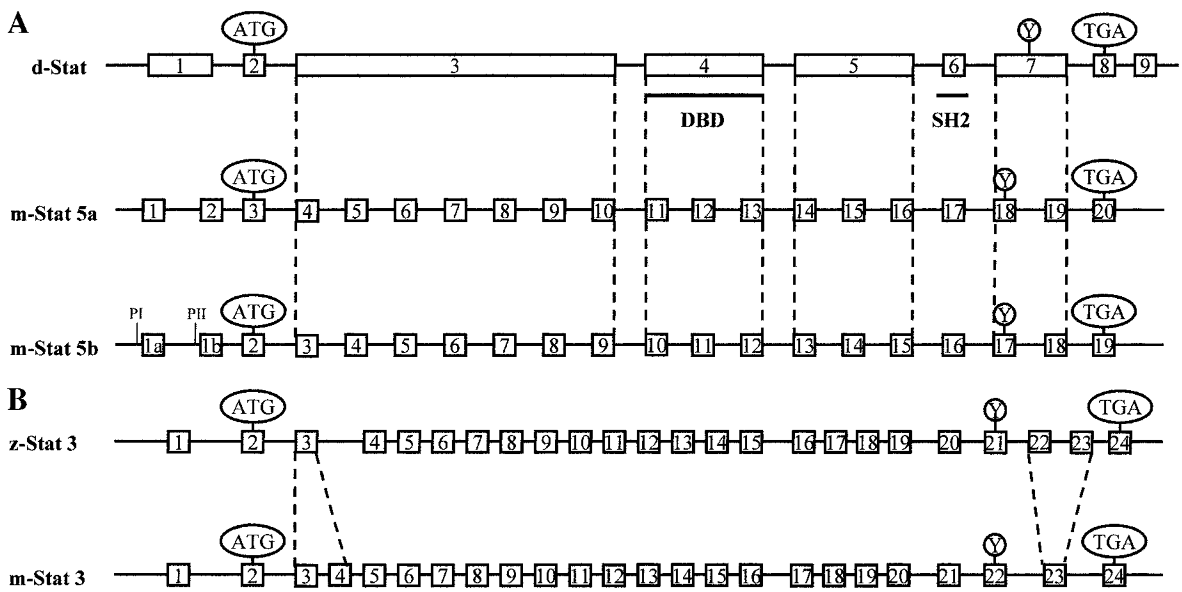

FIG. 2. Structures of m-STATs 3, 5a, and 5b, z-Stat, and d-Stat. (**A**) Comparison of d-Stat and m-STATs 5a and 5b. The drawings depict the intron–exon boundaries (drawings are not to scale). The dashed lines indicate conserved exon boundaries. The d-Stat gene contains nine exons, with the translational initiation codon (ATG) residing in exon 2, the DNA binding domain in exon 4, and the SH2 domain in exon 6; and the tyrosine residue (Y) in exon 7 is phosphorylated by JAK2. The translational stop codon (TGA) is located in exon 8. In the Stat5b gene, promoters PI and PII activate exons 1a and 1b, respectively. (**B**) Comparison of z-Stat3 and m-Stat3.

logue gene mDj11 are located further downstream (Fig. 1C). The third new gene, called D11LGP3, is located upstream of the mDj11 gene (Fig. 1C). The fourth new gene, called D11LGP4, is also located downstream of the Stat5b gene. The new genes were initially identified through a sequence match of Est clones with genomic sequences. Subsequently, full-length mRNAs corresponding to these genes were isolated. While the mouse Ptrf gene is 12 kb in size and consists of 2 exons, the corresponding human gene (TTF-I) consists of only 1 exon. The mouse Hcrt gene is 1.2 kb in size, consists of 2 exons, and encodes the precursor of hypocretin neuropeptides, which stimulate food consumption and regulate food behavior. The human homologue also consists of 2 exons. The mouse GCN5 gene is 7 kb in size and consists of 18 exons. It encodes a histone acetyltransferase directly linking chromatin modification to transcriptional regulation. The Cnp1 gene is 6.8 kb in size, consists of 4 exons, and encodes a phosphodiesterase. The mDj11 gene is 37 kb in size and consists of 14 exons. The function of the encoded protein is not clear, but it may be a chaperone. The homologous human gene TTC2 consists of at least 2 exons.

The mouse Stat5a gene is 30 kb and contains 20 exons with the translational initiation codon located in exon 3 and the translational stop codon in exon 20 (Fig. 2A). The mouse Stat5b gene spans more than 50 kb and contains 20 exons with the translational initiation codon located in exon 2 and the translational stop codon in exon 19 (Fig. 2A). The mouse Stat3 gene has a size of more than 30 kb and consists of 24 exons, with the translational initiation codon located in exon 2 and the translational stop codon in exon 24 (Fig. 2B). Analysis of the 5′UTRs of several mouse Stat5a and Stat3 cDNAs suggested that each gene contains a single promoter. In contrast, mouse Stat5b has two promoters. Promoters I and II are located upstream of exon 1a (88 bp) and exon 1b (463 bp), respectively (Fig. 2A), and are separated by more than 20 kb. Although the genes encoding mouse Stats 5a and 5b have distinct promoters and 5′UTRs, the structure of the protein coding exons is conserved. Using the draft sequence of the human genome, we found the intron–exon boundaries between human and mouse Stats and the position and orientation of Ptrf, Stat3, Stat5a, Stat5b, D11LGP1, and Hcrt to be identical.

The Zebrafish Stat3 Locus

We isolated a BAC clone containing exons 13–24 of the zebrafish Stat3 gene and established its sequence (GenBank Accession No. AF322857). The remaining intron-exon junctions were determined by PCR from zebrafish genomic DNA, followed by sequence analysis. The z-Stat3 gene consists of 24 exons with the translational start and stop codons located in exon 2 and 24, respectively (Figure 2B). A comparison between m-Stat3 and z-Stat3 demonstrates that the presence of an additional intron in the mouse is the result of an intron insertion within exon3 of the z-Stat3 gene (Figure 2B). Exon 24 of both mouse and zebrafish contains the translational stop codon and respective 3’UTRs. The 18 additional amino acids in z-Stat3 are the results of a T to A transition in the stop codon. Exon 22 and 23 in z-Stat3 gene correspond with exon 23 in the m-Stat3 gene.

The Mouse Stat3/5 Locus—Gene Expression

We measured steady-state levels of mRNAs for the genes described above in testis, kidney, muscle, liver,

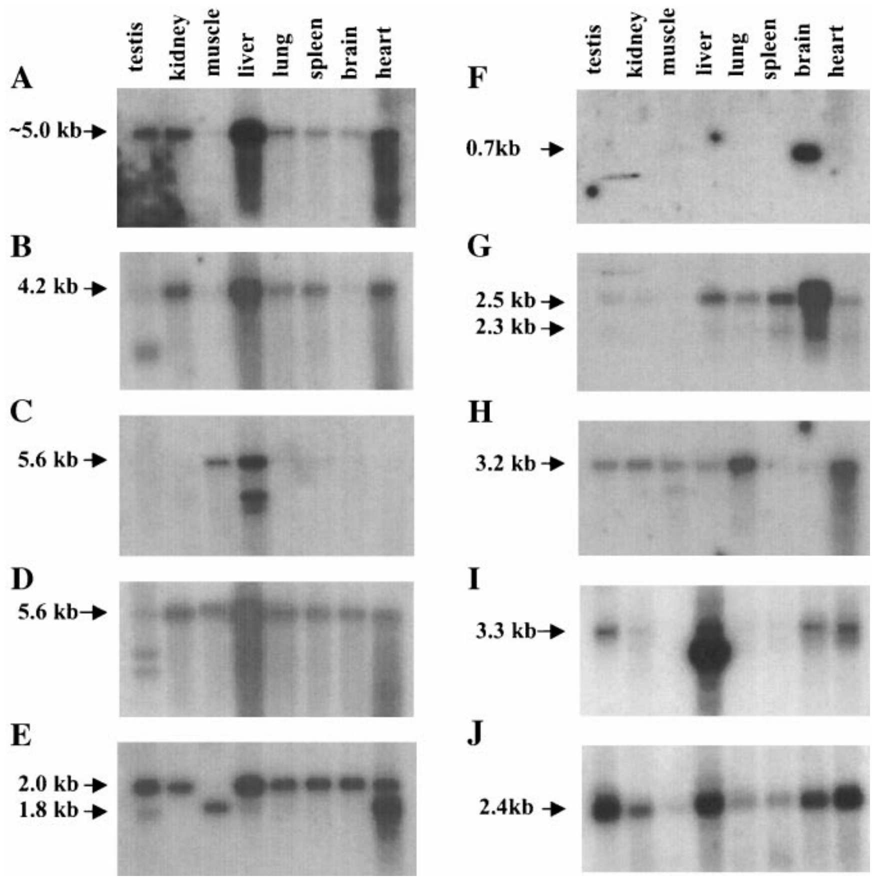

FIG. 3. Expression of genes from the mouse Stat3/5 locus. Steady-state levels of the mRNAs encoded by the eight known genes in this locus were determined in testis, kidney, muscle, liver, lung, spleen, brain, and heart. (**A**) Stat3 mRNA, (**B**) Stat 5a mRNA, (**C**) mRNA from the Stat 5b PII promoter, (**D**) mRNA from the Stat 5b PI and PII promoters, (**E**) β-actin mRNA, (**F**) Hcrt mRNA, (**G**) Cnp1 mRNA, (**H**) Ptrf mRNA, (**I**) GCN5 mRNA, and (**J**) mDj11 mRNA. The probes for Stat3 and Stat5a were from the 3′UTRs. For Stat 5b PII, the probe was from the 5′UTR and recognized the transcript from promoter II. For Stat 5b PI and PII, the probe was from the 3′UTR and recognized the transcripts from promoters I and II.

lung, spleen, brain, and heart (Fig. 3). Stat3 mRNA levels were highest in liver and heart, intermediate in lung, spleen, brain testis, and kidney, and lowest in muscle (Fig. 3A). Similarly, Stat5a mRNA levels were highest in liver and heart, followed by kidney (Fig. 3B). Two Stat5b cDNAs have been isolated that differ in their 5′UTRs (Liu *et al.*, 1995; Mui *et al.*, 1995), which is the result of two distinct promoters. Using a Stat5b 5′UTR probe specific for promoter II, we identified the highest expression in liver followed by muscle (Fig. 3C). Little or no expression from this promoter was detected in other tissues. Since the promoter I-specific 5′UTR was too short and too purine-rich to yield a specific probe, we used a probe from the 3′UTR, which detects transcripts derived from both promoters. While expression was highest in liver, lower levels of mRNA were detected in all other tissues, with the exception of testis, which exhibited a smaller transcript (Fig. 3D). Actin mRNA served as a loading control (Fig. 3E).

We further investigated the expression pattern of the genes flanking the Stat locus. While Hcrt mRNA was restricted to brain (Fig. 3F), Cnp1 mRNA levels were highest in brain, followed by liver, lung, spleen, and heart (Fig. 3G). Ptrf mRNA was highest in lung and heart, and equal but lower levels were seen in all other tissues (Fig. 3H). GCN5 mRNA was highest in liver, followed by heart, brain, and testis (Fig. 3I). RNA from the mDj11 gene was detected in all tissues with testis, liver, heart, and brain having the highest levels of expression (Fig. 3J).

DISCUSSION

Based on comparisons of the protein sequences from d-Stat, m-Stat3, m-Stat5a, m-Stat5b, and z-Stat3 and the exon structures from the corresponding genes, we suggest that these Stat genes have evolved from a common primordial gene consisting of as few as four exons (Fig. 4). We also provide evidence that d-Stat has its origin in the Stat5 lineage. Furthermore we provide evidence that the duplication event leading to the Stat5a and Stat5b genes resulted in the acquisition of an additional, tissue-restricted promoter in the Stat5b gene.

In the mouse, the genes encoding STATs 3, 5a, and 5b are located next to one another on chromosome 11 (60.5 cM). The proteins exhibit an overall identity of 27%, suggesting that the corresponding genes evolved after a duplication event of a primordial gene. Based on the positions of exons in the different Stat genes from

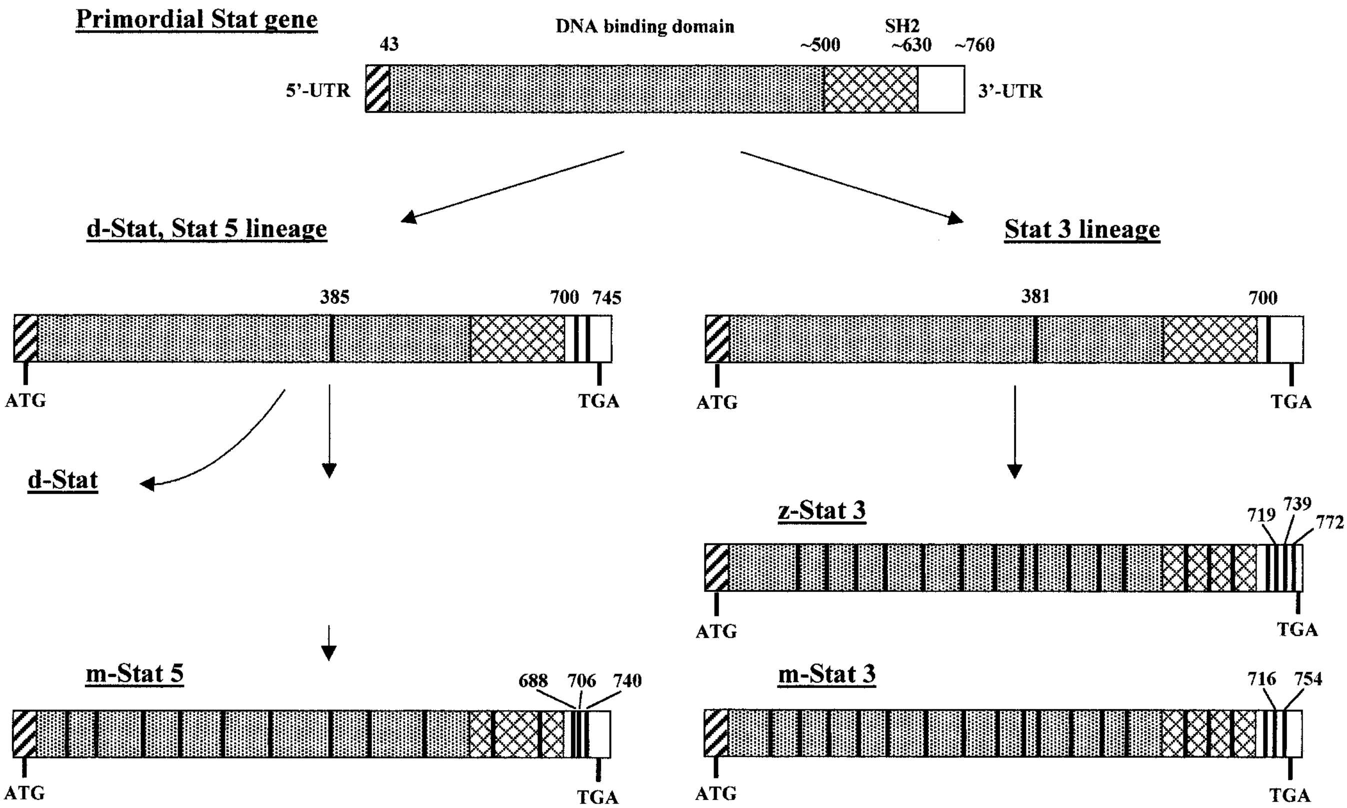

FIG. 4. Evolution model of STATs 3 and 5. We propose that the primordial Stat gene has four exons: two noncoding exons and two protein-coding exons. A duplication of this primordial gene led to the Stat3 and Stat5 lineages. The two lineages are characterized by the insertion of introns at specific sites. Based on the intron–exon junctions, d-Stat is part of the Stat5 lineage. M-Stat5a and m-Stat3 have identical exon structures. Z-Stat3 and m-Stat3 have 24 exons. The additional exons were generated through the insertion of introns in preexisting exons. The striped rectangle represents the 5′UTR, and the open rectangle represents the C-terminus and the 3′UTR. The dotted rectangle depicts the part of Stats that contains the DNA binding domain, and the cross-hatched rectangle depicts the part that contains the SH2 domain. Solid bars indicate introns, and selected insertion sites are numbered. ATG, translational start codon; TGA, translational stop codon; d-Stat, *Drosophila* Stat; z-Stat3, zebrafish Stat3; m-Stat3, mouse Stat3; m-Stat5, mouse Stat5.

Drosophila, mouse, and zebrafish, it is now possible to reconstruct part of the evolutionary trail. We suggest that the primordial Stat gene consisted of four exons, two encoding the 5′ and 3′UTRs and two protein-coding exons (Fig. 4). Upon a duplication event, the Stat3 and Stat5 lineages emerged, which are characterized by specific insertions of intron sequences. Distinct features of the z-Stat3 and m-Stat3 genes suggest that d-Stat evolved from the Stat5 lineage and not the Stat3 lineage. First, while the overall amino acid identity between d-Stat and m-Stat5 is 28%, d-Stat exhibits only 17 and 21% identity with m-Stat3 and z-Stat3, respectively. Second, while exon 4 of the putative primordial gene acquired one intron in z-Stat3 at amino acid 700, d-Stat and m-Stat5 have acquired an additional intron at around position 745 (Fig. 4). We further propose that primordial exons seen in d-Stat acquired additional intron sequences as the Stat5 locus evolved in vertebrates. This is supported by our findings that several of the intron–exon junctions in the d-Stat gene are conserved in m-Stat5 (Fig. 2A). Exon 3 in d-Stat is divided into seven exons in m-Stat5, with the outside boundaries being conserved (Fig. 2A). Exon 4 in d-Stat, which encodes the DNA binding domain, has acquired two additional introns in Stat5a and 5b (Fig. 2A). Exons 5 and 7 in the d-Stat gene are represented in the m-Stat5 genes by three and two exons, respectively (Fig. 2A).

A comparison of the z-Stat3 and m-Stat3 genes provided insight into the order in which introns were acquired (Figs. 2B and 4). Although both Stats 3 contain 24 exons, an additional exon in the mouse gene is the result of intron insertions after mammals diverged from fish. An additional exon in the zebrafish gene is also the result of intron insertion during its evolution. Since d-Stat evolved in the Stat5 lineage, it is possible that *Drosophila* contains a second, more Stat3-like Stat gene. Stat-like activity, which is different from d-Stat, has been reported in *Drosophila* (Zeidler *et al.*, 2000), but no corresponding gene has been identified. Thus it appears that *Drosophila* has lost the Stat gene from this lineage.

The duplication of a postulated Stat 3/5 primordial gene maintained the integrity and specificity of the transcriptional control regions of its descendants. The expression pattern of both Stat5a and Stat3 mRNA is similar in the eight tissues tested and highest in liver. In contrast, the duplication that led to the formation of the Stat5b gene had regulatory consequences in that a new promoter (promoter II) was acquired that is active in liver, muscle, and mammary tissue (Liu *et al.*, 1995). Based on the comparison of 300 bp upstream of the transcriptional start sites, the Stat5a gene promoter is 21% identical with the Stat5b promoter I, but only 6% identical with the Stat5b promoter II.

Walter Gilbert (1978) introduced the concept of exon shuffling as a means to build large numbers of unique proteins through the combination of individual domains encoded by individual exons. However, each of the DNA binding and SH2 domains in m-Stat5 and m-Stat3 is encoded by several exons. The structure of the older *Drosophila* gene indicates that the functional domains were originally encoded in one distinct exon. It is not clear why, during evolution, these larger exons acquired introns during evolution. While the d-Stat exons range in size between 200 and >1000 bp, the m-Stat5 exons range between 90 and 200 bp, and in Stat3 exons are as small as 27 bp. It is possible that the introduction of introns is a random process and the result of mobile elements (Eickbush, 2000). In contrast the 3′UTRs of d-Stat and Stats 3, 5a, and 5b are 1 to 2.5 kb in size. There may be a positive evolutionary pressure to create short coding exons, which facilitates the generation of new proteins. Based on our work, we can demonstrate that a dominant negative form of Stat5a (Kazansky *et al.*, 1995) lacks its C-terminus because the splicing between exon 19 and exon 20 did not occur and a translational stop codon in intron 19 was used. However, the introduction of additional introns during evolution from *Drosophila* to mammals may not be a general one. Recently the SNAP-23 genes from mouse and *Drosophila* have been sequenced (Vaidyanathan and Roche, 2000), and their exon structures are conserved.

The appearance of multicellular organisms created the need for coordinated cell–cell communications in response to extracellular cues. The Jak/Stat pathway is central in conveying many extracellular cytokine signals to the nucleus where they elicit transcriptional responses that lead to cell proliferation, differentiation, survival, and possibly death. Successive gene duplication events resulted in seven mammalian Stat proteins, each of which serves distinct cytokine pathways. The conserved structures of d-Stat, m-Stat5a, and m-Stat5b reflect their related and overlapping functions in hematopoiesis (Harrison *et al.*, 1995; Socolovsky *et al.*, 1999). Based on the overall sequence similarities and exon structures, we propose that m-Stats 3 and 5 are derived from a primordial Stat gene. The zebrafish (*Danio rerio*) genome contains a Stat3 gene, but a Stat5 gene has not been isolated. Since Stat5 is conserved between *Drosophila* and mouse and it controls hematopoiesis, this suggests that it remains to be discovered in zebrafish.

ACKNOWLEDGMENTS

The authors thank the entire genome sequencing group of the Whitehead Institute of Biomedical Research/MIT Center for Genome Research for their efforts in sequencing the BAC clones.

K.M. thanks Professor Hideo Inove (Tokushima University) for his support.

REFERENCES

Chapman, R. S., Laurence, P. C., Tonner, E., Flint, D. J., Selbert, S., Takeda, K., Akira, S., Clarke, A. R., and Watson, C. J. (1999). Suppression of epithelial apopotsis and delayed mammary gland involution in mice with a conditional knockout of Stat3. *Genes Dev.* **13**(19): 2604–2616.

Chatterjee-Kishore, M., van den Akker, F., and Stark, G. R. (2000). Association of STATs with relatives and friends. *Trends Cell Biol.* **10**(3): 106–111.

Eickbush, T. H. (2000). Molecular biology. Introns gain ground. *Nature* **404**(6781): 940–943.

Gilbert, W. (1978). Why genes in pieces? *Nature* **271**(5645): 501.

Hanratty, W. P., and Dearolf, C. R. (1993). The *Drosophila* Tumorous-lethal hematopoietic oncogene is a dominant mutation in the hopscotch locus. *Mol. Gen. Genet.* **238**(1–2): 33–37.

Harrison, D. A., Binari, R., Nahreini, T. S., Gilman, M., and Perrimon, N. (1995). Activation of a *Drosophila* Janus kinase (JAK) causes hematopoietic neoplasia and developmental defects. *EMBO J.* **14**(12): 2857–2865.

Hou, X. S., Melnick, M. B., and Perrimon, N. (1996). Marelle acts downstream of the *Drosophila* HOP/JAK kinase and encodes a protein similar to the mammalian STATs. *Cell* **84**(3): 411–419.

Humphreys, R. C., and Hennighausen, L. (1999). Signal transducer and activator of transcription 5a influences mammary epithelial cell survival and tumorigenesis. *Cell Growth Differ.* **10**(10): 685–694.

Kazansky, A. V., Raught, B., Lindsey, S. M., Wang, Y. F., and Rosen, J. M. (1995). Regulation of mammary gland factor/Stat5a during mammary gland development. *Mol. Endocrinol.* **9**(11): 1598–1609.

Leonard, W. J., and O'Shea, J. J. (1998). Jaks and STATs: Biological implications. *Annu. Rev. Immunol.* **16:** 293–322.

Liu, X., Robinson, G. W., Gouilleux, F., Groner, B., and Hennighausen, L. (1995). Cloning and expression of Stat5 and an additional homologue (Stat5b) involved in prolactin signal transduction in mouse mammary tissue. *Proc. Natl. Acad. Sci. USA* **92**(19): 8831–8835.

Liu, X., Robinson, G. W., Wagner, K. U., Garrett, L., Wynshaw-Boris, A., and Hennighausen, L. (1997). Stat5a is mandatory for adult mammary gland development and lactogenesis. *Genes Dev.* **11**(2): 179–186.

Mui, A. L., Wakao, H., O'Farrell, A. M., Harada, N., and Miyajima, A. (1995). Interleukin-3, granulocyte-macrophage colony stimulating factor and interleukin-5 transduce signals through two STAT5 homologs. *EMBO J.* **14**(6): 1166–1175.

Oates, A. C., Wollberg, P., Pratt, S. J., Paw, B. H., Johnson, S. L., Ho, R. K., Postlethwait, J. H., Zon, L. I., and Wilks, A. F. (1999). Zebrafish stat3 is expressed in restricted tissues during embryogenesis and stat1 rescues cytokine signaling in a STAT1-deficient human cell line. *Dev. Dyn.* **215**(4): 352–370.

Socolovsky, M., Fallon, A. E., Wang, S., Brugnara, C., and Lodish, H. F. (1999). Fetal anemia and apoptosis of red cell progenitors in Stat5a-/-5b-/- mice: A direct role for Stat5 in Bcl-X(L) induction. *Cell* **98**(2): 181–191.

Takeda, K., and Akira, S. (2000). STAT family of transcription factors in cytokine-mediated biological responses. *Cytokine Growth Factor Rev.* **11**(3): 199–207.

Vaidyanathan, V. V., and Roche, P. A. (2000). Structure and chromosomal localization of the mouse SNAP-23 gene. *Gene* **247**(1–2): 181–189.

Zeidler, P. M., Bach, A. E., and Perrimon, N. (2000). The roles of *Drosophila* JAK/STAT pathway. *Oncogene* **19:** 2598–2606.